

A UML-based Meta-Framework for System Design in Public Health Informatics

Anna O. Orlova, PhD
Bloomberg School of Public Health
Baltimore, MD

Harold Lehmann, MD PhD
Johns Hopkins School of Medicine
Baltimore MD

Abstract

The National Agenda for Public Health Informatics calls for standards in data and knowledge representation within public health, which requires a multi-level framework that links all aspects of public health.

Method: *The literature of public health informatics and public health informatics application were reviewed. A UML-based systems analysis was performed. Face validity of results was evaluated in analyzing the public health domain of lead poisoning.*

Results: *The core class of the UML-based system of public health is the Public Health Domain, which is associated with multiple Problems, for which Actors provide Perspectives. Actors take Actions that define, generate, utilize and/or evaluate Data Sources. The life cycle of the domain is a sequence of activities attributed to its problems that spirals through multiple iterations and realizations within a domain.*

Conclusion: *The proposed Public Health Informatics Meta-Framework broadens efforts in applying informatics principles to the field of public health.*

Introduction

Public health informatics is defined as the application of information science and technology to public health practice and research.¹ One contribution of informatics to clinical care has been the adoption of a variety of communication and terminological standards to information systems that, once thought independent, now need to interact. The applications of these informatics principles to public health were formulated as a National Agenda for Public Health Informatics of the AMIA 2001 Spring Congress.² To create similar standards for public health requires an analysis of public health in terms of its content and practice.

Public health is defined by public health researchers as a discipline that addresses “diverse range of problems and, consequently, [involves] the broad scope of activities.”³ Public health is a multi-discipline field of endeavor traditionally represented by (but not limited to) the following areas: epidemiology, environmental health sciences, occupational health sciences, behavioral sciences, health care management and health policy development.³ Generation of public health

information involves clinical data as well as data collected in other fields, such as environmental,² large-scale demographic, and geographic data. Therefore, public health information infrastructure is “broader than [what] is traditionally addressed by medical informatics.”¹

In 2000 the CDC developed a Public Health Conceptual Data Model that includes 4 subject areas (health related activities; locations; materials and parties) and contains 29 classes.⁴ This model focuses on interactions among public health and health care systems. But “public health as a discipline encompasses an amalgam of science, action, research, policy, advocacy and government.”³ Therefore, efforts are needed to describe interactions of public health beyond clinical encounters.

The goal of this paper is to present a Public Health Informatics Meta-Framework (PHIMF) for understanding the complexity and organization of public health information at a high level of abstraction. This meta-framework should be thought of as part of the effort needed to create reference information models (RIMs)⁵ in public health, by specifying commonalities and regularities among public health domains, and by defining elements that comprise public health data and knowledge. Using RIMs enables interoperability and reuse of data.

We base our framework on the Unified Modeling Language (UML)⁶ for a number of reasons. It is an emerging standard at the graphical and conceptual levels. The resulting framework can therefore be shared among researchers building RIMs. The result of the UML specification process is not just a document that uses commonly understood iconography, but object class hierarchies and even code stubs that can be used to construct operating systems.

Method

A review of 441 publications (January 1995 through December 2000)⁷ and of the NLM’s MeSH hierarchy was performed to understand the current status of research and system development related to public health informatics. The public health domain² was used as the core object of concern. The life cycle of the public health domain, in terms of data and knowledge generation was analyzed based on a

“Problem → Response” public health approach developed by CDC National Center for Injury Prevention and Control.³

We further tested the model by applying it to describe a public health domain of lead-poisoning -- one of the main public health problems of environmental etiology in the US and worldwide.⁸⁻¹⁰ An iterative UML process was used, by building Use Cases, State Sequence, and Class Diagrams. Poseidon for UML/Community Edition 1.1 was used.¹¹

Results

Fig. 1 provides the high-level Use Case diagram for public health informatics. Fig. 2 shows the class hierarchy that underlies PHIMF. The core class in the model is the Public Health Domain (see Fig. 1), *e.g.*, Lead Poisoning or Trauma/Injury. Domain(s) represents an area of public health concern, independent of the methods used to manage them. A Domain (core class) may be associated with other Domains (not shown in Fig. 2), *e.g.*, Lead Poisoning is an Environmental Illness. A Domain may include many Problems (see Fig. 2).

For lead poisoning, examples of Problems include: (1) understanding lead poisoning and identifying biological indicators of lead toxicity (*e.g.*, blood lead levels)¹²; (2) understanding and addressing long-term health and social effects of lead poisoning (*e.g.*, learning disabilities, violent behavior)¹³; (3) understanding routes of environmental lead exposure (*e.g.*, dust, soil, water)¹⁴; (4) developing and implementing strategies of primary prevention by eliminating and controlling sources of lead exposure (*e.g.*, leaded gasoline, lead-based paint);¹⁴ and (5) developing and implementing secondary prevention measures by: identifying groups of population at risk (*e.g.*, young children, pregnant women, workers) through blood lead screening; providing medical treatment to lead poisoned patients; and conducting follow-up case management activities to minimize lead exposure.^{16, 17}

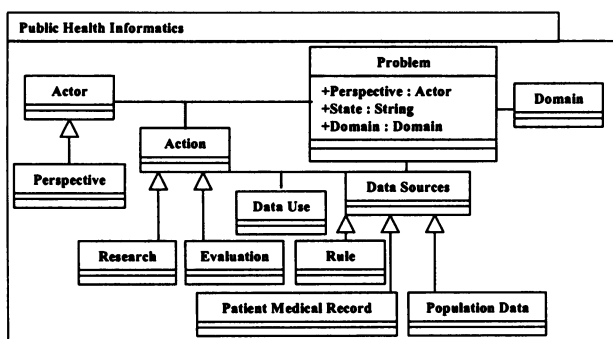


Figure 2. Class Hierarchy

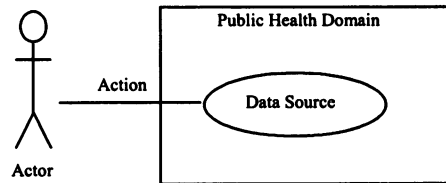


Figure 1. High-level Use Case Diagram for Public Health Informatics

An association of the Problem class with an Actor class is an Action class, *i.e.*, an Action involved in addressing (solving) the Problem (see Fig. 2). Actions are performed by Actors, the people or agents who play different roles in addressing a Problem. In lead poisoning, Actors include Policy Makers, Service Providers, Researchers, the Public, Clinicians, and Educators. Thus, considering the Problem of lead poisoning primary prevention, Policy Makers would want to solve this problem to improve population health by eliminating sources of lead exposure (*e.g.*, leaded gasoline, lead-based paint), while Service Providers (*e.g.*, local health departments) would want to solve this problem to identify and provide services to people at risk of lead poisoning.

To address the Problem, Actors take Action(s), which, among other things, involves creation and use of Data Sources (see Fig. 1 and 2). In public health, such Actions include Surveillance, Risk Assessment, etc. Data generated within these actions include Registries, Research Databases, and so on. Lead poisoning primary prevention, for instance, under the Maryland Reduction of Lead Risk in Housing Law¹⁸ involves (1) registration of certain rental properties, likely to contain lead-based paint, in the Maryland's Department of the Environment's (MDE) Rental Registry; and (2) lead risk reduction treatments certified and recorded in the MDE Oversight database. Lead poisoning secondary prevention includes annual blood lead screening of children under age of 6, and the results of these tests are collected in the state Maryland Blood Lead Registry.¹⁹ Blood lead registry data in turn are used for case management to understand sources of lead exposure in the environment of the child with elevated blood lead level and to eliminated these sources, *e.g.*, by relocation of family with lead poisoned child to a lead-free house.²⁰

A life of a public health problem can be described in the Sequence Diagram (Fig. 3) using and extending the “Problem → Response” framework.³ The transitions presented on the diagram are not exhaustive. The Sequence diagram shows Actions within each state in time. Boxes represent state. Lines

with arrows show how generated data feed into the next class or state. Each next state utilizes data from the previous state, generates new data that help understand previous data and therefore improves overall knowledge about the problem and the ways of addressing the Problem.

The multiple transitions between states (Problem Identified, Problem Characterized, Problem Managed, and Problem Evaluated) show how different Actions can result in changing a Problem's state. A particular pathway through the diagram leads to, or defines, the information system of an individual public health problem.

The process of making a Problem Identified in public health is the more poorly formalized step of the process, as it is in medicine.²¹ It may result from informal, anecdotal data and/or from a formal, systematic description of health problem. However, until a Problem is documented in a Data Source, it does not exist. For instance, lead poisoning, although known for centuries, did not reach public-health awareness until clinical data were published.⁹ Next is the question of etiology and causality. Answers are suggested by the analysis of similar cases from population-based Data Sources as well as data gathered through individual patient medical records (another Data Source) and population based data.

Making a Problem Characterized (see Fig. 3) involves systematic investigation of the potential reasons that caused the Problem to appear, based on the best available research methods. Epidemiological Surveillance is a class of Actions used to characterize population-based data about a public health problem (e.g., lead poisoning data, injury data, etc.). Risk assessment is a class of Actions used (not limited) to characterize public health problem of environmental etiology, ²² e.g., lead poisoning.²³ Research results may reside in Research Databases or in Journal Articles and Books (further Data Sources).

Getting a Problem Managed is accomplished by developing guidelines, and by promulgating bureaucratic rules and regulations (more Data Sources) for how to manage the problem. In lead poisoning, guidelines for blood lead screening among young children were developed by CDC and state Blood Lead Registries were established¹⁵; and health-

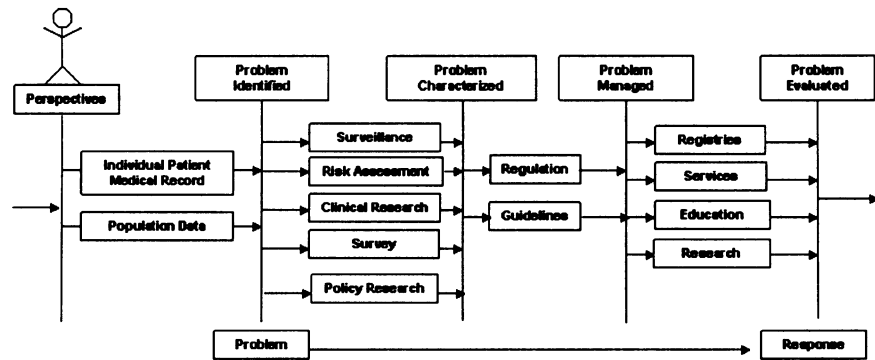


Figure 3. Sequence Diagram for Problem states

based standards for lead in the environment have been instituted by US Environmental Protection Agency (EPA).²⁴ EPA's Data Sources here are currently report-based and some are online, but not queryable.

Getting a Problem Managed (see Fig. 3) also involves developing Services to ensure that population in need can be identified and can have adequate access to health care. For lead poisoning, this has included establishing specialized Lead Clinics (e.g., Kennedy Krieger Institute's Lead Clinic in Baltimore, Maryland) that resulted in further clinical records as Data Sources. Environmental sources of lead exposure have been identified and controlled by banning some hazardous materials (e.g., leaded gasoline and lead-based paint),⁹ and/or by implementing lead risk reduction strategies (e.g., lead abatement) to reduce and/or to eliminate sources of lead exposure.^{17, 25} The associated Data Sources currently exist in Journal Articles, Books, and Technical Reports.

As the final example of getting a Problem Managed, educational campaigns are developed to target population at risk and general public to prevent health problem. For instance, EPA developed educational pamphlets for lead poisoning prevention.²⁶ Local health departments have community educators to deliver educational materials to the public, e.g., Baltimore City Health Department's Lead Program.²⁷ EPA educational material are available on-line.²⁸

The Data Sources generated by Actions that result in a Problem Managed may be used for evaluation of the effectiveness of the problem management interventions, resulting in a Problem Evaluated state (see Fig. 3). Further improvement of problem management strategies through further actions (i.e.,

Table 1. Realizations of the class hierarchy

Domains	Actors/Perspectives	Problems	Actions	Data Sources
Infectious diseases	Elected official	Understanding	Surveillance	Registry
Injury/Trauma	Policy maker	Primary prevention	Screening	Patient medical record
Bioterrorism	Health department	Secondary prevention	Survey	Governmental regulation
Congenital disease	Researcher	Tertiary prevention	Regulation	Guideline
Consumer product safety	Private sector stakeholder		Education	Research database
Environmental hazards	Clinician		Evaluation	Decision support system
Drug abuse	Educator		Risk assessment	Peer-reviewed and non-peer-reviewed literature
Mental health	Citizen		Policy research	

policy development, new research, education and services) may lead to elimination of the problem and to prevention it from re-occurrence.

More likely, new Problems are identified, and the cycle begins anew. Therefore, it should be noted that in the Sequence Diagram (Fig. 3), the Problem Evaluated state associates with the starting point of the diagram as well, and, the proper geometry would be a spiral, as time moves forward.

To summarize these results, Table 1 provides a non-exhaustive list of examples (realizations) of each of the major classes in PHIMF. To generalize our analysis, we would expect that the classes, viewed as dimensions, should be independent of each other. We find this to be the case in Table 1: a mix of items, one from each column, leads to a meaningful relationship in public health. Following the example of lead poisoning, any of the 7 listed Actors may be involved in addressing problems within this public health domain. Registries of residential dust samples, blood-lead registries, Geographical Information Systems applied to environmental data, clinical medical records, guidelines and regulations, and rules are needed to address lead poisoning prevention.

Discussion

We present a Public Health Informatics Meta-Framework for modeling frameworks of components of public health information systems. A meta-framework is important because, since public health covers wider levels of organization than clinical medicine, so frameworks for specific Domains may be unwieldy and will need to be separated.

PHIMF separates concepts of public health that may currently be perceived as lumped together. For instance, the current MeSH hierarchy includes as immediate descendants under the term "Public Health" Domains, like "Environmental Pollution"; Problems, like "Carrier State"; and Actions, like "Epidemiological Methods."

PHIMF suggests that some current conceptions must be modified. For instance, one focus of attention in

epidemiology is defining burden of illness in a population. "Illness" becomes a Domain with Problems, while epidemiological methods become associated with Actions. So the traditional "domain" of epidemiology gets split into components that lie in different places in the meta-framework.

PHIMF also separates data management functions from problem-solving functions. Data Sources must therefore be used for multiple purposes, either for transitions between different states, or by different Actions or by different Problems altogether. This reuse implies that Data Sources must be designed to be reused, or must be designed to communicate with other Data Sources in the State Sequence model (Fig. 3). PHIMF may serve the basis of federating databases in public health.²⁹

Other distinctions include attitude towards Data Sources. While most discussion in public health has revolved about data systems, our analysis points to viewing other repositories as active participants: books, guidelines, governmental regulations, and reports among them. For instance, our meta-framework can be used for creating filters for PubMed that are appropriate for public health readers, just as the currently available Clinical Query feature provides filters for clinical questions.

UML is increasingly used in depicting the types of frameworks discussed here. For instance, Winter and colleagues³⁰ present multiple views of hospital-information-systems architectures. Using their labeling, we have focused on the Domain Layer (Fig. 2), leaving Logical Tool and Physical Tool Layers for further analysis. It is the working hypothesis of our effort that, by separating the Domain issues from Data Source (Logical Tool) issues, regularities (stereotypes) in schemas can be defined in, say, a registry system designed for lead poisoning may be used in other Domains outside lead poisoning. We are currently exploring this hypothesis in collaboration with researchers in autism research.³¹

There are other approaches to the interoperability problem that we are addressing. One extreme is to get

agreement at the data level and then to focus on data collection forms and communication, much as HL7 did before RIM 3.0. The other extreme is to rely on meta-query brokering. In this solution, data source owners translate their schemas into a high-level schema,^{32,33} or have the meta-brokering system perform the translation.^{34,35} Our effort would work more closely with the latter approach, with PHIMF providing just the sort of "thin domain model" needed by the meta-query process.

Our efforts are concordant with National Agenda for Public Health Informatics.² Specifically, they apply to the following Agenda's items: Architecture and infrastructure (A-5, A-9, A-10); Standards and vocabulary (S-3, S-5, S-8, S-9, S-10); Research, evaluation and best practices (R-10, R-11, R-12); Training and workforce (T-10, T-13).

Acknowledgments

Thanks to Nancy Roderer for formative suggestions, to the Welch Medical Library staff for helping with the literature review, and to Paul Law for reviewing an earlier draft.

References

1. Friede A, McDonal M, Blum H. Public health informatics: how information-age technology can strengthen public health. *Ann Rev Public Health* 1995;16:239-52.
2. Yasnof W, Overhage J, Humphrey B, LaVenture M. A national agenda for public health informatics. *J Am Med Inf Ass* 2001;8(6):535-45.
3. Koo D, O'Carroll P, Laventure M. Public health 101 for informaticians. *J Am Med Inf Ass* 2001;8(6):585-97.
4. Centers for Disease Control and Prevention (CDC). Public Health Conceptual Data Model: Premiere Edition. Atlanta, GA; 2000. <<http://www.cdc.gov/od/hissb/docs/PHCDM%20Table%20Contents.pdf>>
5. HL7 Standards: HL7; 2002. <<http://www.hl7.org/Library/standards.cfm>>
6. Rumbaugh J, Jacobson I, Booch G. The Unified Modeling Language reference manual. Reading, MA: Addison-Wesley; 1999.
7. Current bibliographies in medicine. Public health informatics. January 1996 through December 2000. Washington, DC: National Institutes of Health. National Library of Medicine; 2001. Report No.: 2001-2.
8. Centers for Disease Control and Prevention (CDC). Preventing lead poisoning in young children. A statement by the Centers for Disease Control. Atlanta, GA; 1991.
9. Silbergeld E. The international dimensions of lead exposure. *Int J Occup Environ Health* 1995;1(4):336-348.
10. Thomas V, Orlova A. Soviet and post-Soviet environmental management: lessons from a case study on lead pollution. *Ambio* 2001;30(2):104-111.
11. Gentleware.com. Poseidon for UML. In. Community Edition 1.1 ed; 2001. <<http://www.gentleware.com/>>
12. Silbergeld E, Davis D. Role of biomarkers in identifying and understanding environmentally induced disease. *Clin Chem* 1994;40(7 Pt 2):1363-7.
13. Needleman H, A, Bellinger M, Leviton A, Allred E. The long-term effects of exposure to low doses of lead in childhood: An 11-year follow-up report. *N Engl J Med* 1990;322(2):83-88.
14. Agency for Toxic Substances and Disease Registry (ASTDR). Toxicological profile for lead. Atlanta, GA; 1999.
15. Chisolm Jr J, Farfel M. Environmental control and deleading. *Pediatr Ann* 1994;23(11):627-8, 631-3.
16. Centers for Disease Control and Prevention (CDC). Screening young children for lead poisoning: Guidance for state and local public health officials. Atlanta, GA; 1997.
17. Lanphear B, Howard C, Eberly S, Auinger P, Kolassa J, Weitzman M, et al. Primary prevention of childhood lead exposure: A randomized trial of dust control. *Pediatrics* 1999;103(4 Pt 1):772-7.
18. Reduction of Lead Risk in Housing. In: Annotated Code of Maryland; 1994. p. Environment Article 1-404, 6-801 - 6-852, 6-1001 - 6-1005.
19. Keyvan. Maryland Blood Lead Registry. Personal Communication. 1998.
20. Farfel M, Quinn R. A lead-safe family shelter in an urban minority community. *Am J Public Health* 1994;84(8):1338-9.
21. Blois M. Information and medicine : the nature of medical descriptions. Berkeley: University of California; 1984.
22. Burke T. Understanding environmental risk: the role of the laboratory in epidemiology, evaluation, and policy setting. *Clin Chem* 1992;38(8B Pt 2):1519-22.
23. Succop P, Clark S, Tseng C, Bornstein R, Chen M. Evaluation of public housing lead risk assessment data. *Env Geochim & Health* 2001;23(1):1-15.
24. U.S. Environmental Protection Agency (EPA). Lead and lead compounds; lowering of reporting thresholds; Community right-to-know toxic chemical release reporting, Final rule. In: Federal Register; 2001.
25. Sargent J, Dalton M, Demidenko E, Simon P, Klein R. The association between state housing policy and lead poisoning in children. *Am J Public Health* 1999;89:1690-1695.
26. U.S. Environmental Protection Agency (EPA). Protect your family from lead in your home. Washington, D.C; 1995. Report No.: EPA 747-K-94-001.
27. Endres J, Montgomery J, Welch P. Lead poison prevention: a comparative review of brochures. *J Env Health* 2002;64(6):20-5.
28. Lead in Paint, Dust, and Soil. U.S. Environmental Protection Agency (EPA); 2002. <<http://www.epa.gov/lead/leadpbed.htm>> Date last verified, June 24, 2002>
29. Roantree M, Hickey P, Crilly A, Cardiff J, Murphy J. Metadata modelling for healthcare applications in a federated database system. In: Spaniol O, Linnhoff-Popien C, Meyer B, editors. Trends in Distributed Systems: CORBA and Beyond. International Workshop TreDS '96; 1996 October 1-2, 1996; Aachen, Germany; 1996. p. 71-83.
30. Winter A, Brigl B, Wendt T. A UML-based ontology for describing hospital information system architectures. In: MEDINFO 2001; 2001; London, UK: IOS; 2001. p. 778-82.
31. Law P. An introduction to Isaac. Internet Systems for Assessing Autistic Children; 2000. <<http://www.autismtools.org/isaacfront/lawletter.cfm>> Date last verified, Feb 28, 2002>
32. Bouguettaya A, Benatallah B, Elmagarmid A. Interconnecting Heterogeneous Information Systems. Dordrecht: Kluwer; 1998.
33. Elmagarmid A, Rusinkiewicz M, Sheth A, editors. Management of Heterogeneous and Autonomous Database Systems: Morgan Kaufmann; 1999.
34. Silberberg D, Semmel R. Using Relationship Roles to Disambiguate Conceptual-Level Queries. In: Deng Y, Gerken M, editors. Tenth International Conference on Software Engineering and Knowledge Engineering; 1998; Skokie, IL: Knowledge Systems Institute; 1998. p. 327-334.
35. Silberberg D. Uniform and High-Level Intelligent Access To Heterogeneous Information Sources. College Park, MD: University of Maryland; 2002, to appear.